

BEYOND UTILISATION

The Pipeline Illusion: Why a Full Portfolio Signals Overload, Not Ambition

Reflections on capacity, demand and the arithmetic portfolio governance keeps refusing to do



Contents

1. Executive Summary	3
2. The wall of green	3
3. The illusion we reward	4
4. Demand has no natural ceiling	4
5. The arithmetic nobody does	5
6. What queues know that committees do not	5
7. "But idle capacity is waste"	6
8. Why the illusion persists	7
9. From approving to finishing	7
10. The courage of the empty slot	8

The Pipeline Illusion: Why a Full Portfolio Signals Overload, Not Ambition

Portfolio · No. 21

Originally written September 2006 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“The uncomfortable truth is that ambition is demonstrated by what finishes, not by what is approved, and the surest way to finish more is to start less.”

1. Executive Summary

Walk into any portfolio review and the instinct is to admire a full pipeline. A long list of approved initiatives looks like ambition, appetite, an organisation betting on its own future. This essay argues the opposite: that a full portfolio is far more often a symptom of overload than a sign of health, and that the fullness we celebrate is precisely what guarantees that little of it will finish.

The illusion has a simple mechanism. Demand for change has no natural ceiling, while the capacity to deliver it does. Governance that approves faster than it sequences quietly loads the organisation past the point where anything flows, and then mistakes the resulting congestion for productivity. The queue is invisible on the plan; the delay is very visible in the results.

I want to do three things. First, name the illusion and show the arithmetic that governance keeps declining to do. Second, take seriously the strongest defence of the full pipeline — that idle capacity is waste and a deep backlog is prudent — and explain why it is wrong in a system where work must flow, not merely be busy. Third, describe what a portfolio managed for throughput rather than appetite actually looks like: fewer things open, finished sooner, and the uncomfortable discipline of leaving capacity unspent on purpose. The counting we need is not accounting. It is queueing.

2. The wall of green

The quarterly portfolio review begins, as it always does, with the wall. Ninety initiatives, each a row, each a status: a reassuring field of green, a scatter of amber, two red items that everyone has already agreed to watch. The director running the meeting is quietly proud of the density. A year ago the list held sixty items; now it holds ninety. Demand is up, the business is engaged, the change function is trusted with more than ever. Nobody in the room says the obvious thing, because the obvious thing does not fit the mood: almost nothing on this wall is actually moving.

I have sat in that room more times than I can count, and the tell is always the same. Ask not "what is the status of each initiative?" but "how many things did we finish last quarter?" The status wall answers the first question fluently and cannot answer the second at all. When someone finally works it out from the closure log, the number lands with a small shock. Ninety open. Six closed in the year. The portfolio is not a pipeline through which work flows; it is a car park in which work is stored, and the fuller it gets, the less anything

can move.

That is the illusion in miniature. Fullness reads as vitality. A busy portfolio feels like a productive one. And the feeling is not merely wrong — it is exactly backwards, because the very act of filling the portfolio is what stops it delivering.

3. The illusion we reward

It is worth being precise about why the illusion is so durable, because it is not stupidity that sustains it. It is a set of perfectly rational local incentives that add up to a collective error.

Approving an initiative is an act of optimism, and optimism is cheap to grant. The sponsor leaves the committee happy. The strategy is seen to be funded. The change function looks capacious and willing. Every individual approval is defensible; each one clears its own business case, each promises a return, each has an advocate who will remember who blocked it. What no one owns is the aggregate — the simple fact that the ninety approvals, taken together, describe far more work than the organisation can perform in any period that matters.

No single approval overloads a portfolio. Overload is what you get when every approval is judged on its own merits and none is judged against the capacity of the whole.

So the portfolio grows, and its growth is read as a good news story. We have moved, the narrative goes, from doing projects to managing a portfolio — a genuine advance, and one the profession is only now formalising, with portfolio management at last acquiring a published standard of its own. But a great deal of what passes for portfolio management is really portfolio accumulation: a register that records what has been approved, dressed as an instrument that governs what will be done. The two are not the same, and the gap between them is where delivery quietly dies.

4. Demand has no natural ceiling

The deeper reason the pipeline fills is structural, and it is this: the demand for change is effectively unbounded, while the capacity to deliver it is fixed and stubborn.

Every function in an organisation can always name three things it would improve if only it had the resource. New regulation arrives and must be absorbed. A competitor moves and a response is drafted. A new director joins and brings a mandate that becomes four initiatives by the second month. None of this is illegitimate; all of it is real. Demand is a fountain, not a bucket. It refills as fast as you draw from it, and it never signals "enough."

Capacity, by contrast, is embarrassingly finite, and finite in a particular, awkward way. It is not headcount in the abstract — it is the handful of people who can actually do the load-bearing work. In most organisations a small number of individuals sit on the critical path of a disproportionate share of the initiatives: the systems integration lead every project's data migration depends on, the two business architects who understand how the processes truly connect, the finance partner who must sign every benefits case. You can hire generously around them and barely move the constraint, because the constraint was never the average worker. It was the specific, scarce, un-clonable judgement that a third

of the portfolio is quietly queuing for.

When unbounded demand meets a fixed constraint and governance keeps saying yes, the outcome is arithmetic, not misfortune. The system fills to its limit, and then past it, and the surplus does not vanish. It turns into a queue.

5. The arithmetic nobody does

Here is the calculation a portfolio review almost never performs, and it takes ten minutes.

Take the ninety approved initiatives and estimate, however roughly, the demand each places on the genuinely scarce roles over the coming quarter. Sum it. Then take the actual available capacity of those roles in the same quarter – net of leave, of the operational duties they never escape, of the meetings that governance itself imposes. Put the two numbers side by side.

Measure	What the plan assumes	What is actually available
Demand on the critical roles this quarter	~210 person-months, if every approved initiative proceeds as chartered	The same 210 – the plan simply assumes it can be met
Capacity of those roles this quarter	Treated as sufficient, because each project was approved in isolation	~130 person-months, net of operations, leave and governance overhead
Implied position	A balanced, ambitious portfolio	A structural over-commitment of roughly sixty per cent, discovered only in the slippage

The figures are composite, but the shape is one I have seen hold across sectors and across two decades: demand on the constraint running fifty to seventy per cent above what the constraint can supply. And the moment you see it written down, the status wall reinterprets itself. Those green initiatives are not progressing slowly because they are difficult. They are progressing slowly because each one is waiting, most of the time, for a person who is committed to eight other things. The customer-billing replacement that ran sixteen months late did not fail on its technical merits; its lead analyst was named on the charters of seven other initiatives, and for most of any given month she was simply somewhere else. The delay was not a project problem. It was a portfolio problem wearing a project's name.

The reason the arithmetic goes undone is not that it is hard. It is that its answer is unwelcome. To do it honestly is to admit that a third of the approved portfolio cannot start when it is scheduled to start, and that admission implies a decision no one in the room wants to be the author of: something must be stopped.

6. What queues know that committees do not

There is a body of thinking that explains all of this precisely, and it did not come from the change profession. It came from the study of factories, of operations, of any system through which work must flow.

The first lesson is the distinction between utilisation and throughput. Utilisation asks: are our people busy? Throughput asks: how much finished work leaves the system? We assume the two rise together — that a busier organisation is a more productive one — and for a lightly loaded system that is roughly true. But every operations manager who has run a real plant knows the crueller truth: as you push utilisation towards a hundred per cent, the time it takes any single item to get through the system does not rise gently. It climbs, then it soars. A system loaded to ninety-five per cent is not a little slower than one loaded to seventy. It can be several times slower, because at high load every task spends almost all of its life waiting behind other tasks for the same scarce resource.

“A portfolio loaded to the limit is not maximally productive. It is maximally congested, and congestion is the one thing a status report cannot show.”

The second lesson is older still and can be stated almost as a law: the more things you have in progress at once, the longer each one takes to complete. Cut the number of things open concurrently and the average time-to-finish falls, mechanically, without anyone working harder or later. Goldratt taught a generation of manufacturers to manage the constraint and stop measuring local busyness; the same logic, unglamorously, governs a portfolio. The organisation with ninety initiatives open is not ahead of the one with thirty. It is behind, because its work-in-progress is three times as high and its flow, therefore, three times as slow — and it has ninety opportunities to feel busy while finishing less.

None of this is exotic. It is the physics of queues, and it applies to a portfolio review as surely as to a production line. The tragedy is that governance is conducted almost entirely in the language of accounting — cases, returns, budgets, utilisation — and almost never in the language of flow. We count what has been committed. We do not count what is waiting.

7. "But idle capacity is waste"

The strongest objection to everything I have said is not naïve, and it deserves to be met at its strongest rather than caricatured.

It runs like this. Capacity is expensive and perishable; an hour of a scarce architect's time not used today cannot be banked for tomorrow. A deep pipeline is therefore prudent, not reckless: it ensures that whenever a person comes free, there is always valuable work for them to pick up, so the costly resource is never left idle. Slack is waste. A full backlog is simply good stewardship of an expensive asset, and an organisation that deliberately left capacity unspent would be an organisation throwing money away. Better a queue than an idle hand.

I have made a version of this argument myself, and there is a real truth inside it: capacity genuinely is perishable, and gross underloading genuinely is wasteful. An organisation running its critical people at forty per cent is not virtuously managing flow; it is under-investing, and it should take on more.

But the argument fails at the point that matters, because it optimises the wrong thing. It seeks to maximise the utilisation of the resource, when the goal is to maximise the

throughput of the system. Those two objectives diverge sharply near full load, and the divergence is the whole game. Keeping every architect busy at ninety-five per cent feels efficient and is, for the architect. For the organisation it is ruinous, because it is exactly the condition under which every initiative's time-to-finish soars and the cost of delay — real money, foregone benefits, missed windows — accumulates across the entire portfolio at once. The saving from the last five per cent of utilisation is trivial. The cost of the delay it creates, multiplied across ninety initiatives, is enormous. A modest amount of deliberately unspent capacity is not waste. It is the price of flow, and it is far cheaper than the congestion it prevents.

The full backlog, in other words, does not protect against idleness. It manufactures delay, and then hides the bill inside the slippage of everything at once.

8. Why the illusion persists

If the arithmetic is this simple and the physics this well understood, why does the pipeline keep filling? Because every force in the system pushes towards accumulation and almost none pushes towards restraint.

- Approval is easier than sequencing. To approve is to say yes once, to one sponsor, on one case. To sequence is to tell a room full of sponsors that their initiatives are real, funded, and still not starting yet. The first is a pleasant afternoon; the second makes enemies. Governance drifts towards the pleasant option.
- The annual cycle funds everything at once. When budgets are set in a single yearly round, every approved initiative is handed its money in the same January and told, implicitly, to begin. The calendar itself creates a stampede at the start line, with no mechanism to stagger entry as capacity frees up.
- Utilisation is the metric we can see. Busyness is easy to measure and comfortable to report; flow is not. So we manage the number we have — are people fully loaded? — and neglect the number we need — is work actually finishing? We optimise the visible proxy and lose the real objective.
- No one owns the aggregate. Each sponsor owns an initiative. Each delivery lead owns a workstream. The portfolio-level truth — that the sum exceeds the capacity — belongs to no single person with the authority and the incentive to act on it. It is everybody's context and nobody's problem.
- Stopping feels like failure. To cancel an approved initiative reads, in most cultures, as an admission of error rather than an act of discipline. So work is rarely killed; it is merely starved, kept nominally alive on the register while receiving no real capacity — which is the worst of both worlds, carrying all the overhead of an open initiative and delivering none of the benefit.

Put these together and the full pipeline is not an accident or a lapse. It is the equilibrium the system naturally reaches when approval is frictionless, funding is synchronised, busyness is the measure, the aggregate is unowned, and stopping is taboo. Change any one of those and the picture starts to shift.

9. From approving to finishing

What, then, does a portfolio managed for throughput rather than appetite look like? Less crowded, and considerably more honest.

1. Govern the aggregate, not the item. The first duty of a portfolio board is not to judge initiatives one by one but to hold the sum against capacity and refuse to let it exceed the constraint. That means the operative question at the gate is no longer "is this a good initiative?" — most are — but "is this good enough to displace something already in flight, given that we are full?" Approval becomes a comparison, not a blessing.
1. Limit the work in progress on purpose. Set an explicit ceiling on how many initiatives may be genuinely active at once — a number derived from the capacity of the scarce roles, not from the appetite of the sponsors. New work does not start because it was approved; it starts when something finishes and a slot opens. The register may hold a hundred ideas. The active list holds only what the constraint can actually feed.
1. Finish before you start. Bias the whole system towards completion. When capacity frees, the default is to pour it into the initiatives closest to done, not to sprinkle it across new starts. A portfolio that religiously finishes its nearest-complete work turns its work-in-progress over faster, banks benefits sooner, and — because flow time falls as WIP falls — completes more in a year despite starting less.
1. Stagger entry against real capacity. Break the tyranny of the annual start line. Let initiatives enter the active portfolio as slots open through the year, so the constraint is fed steadily rather than flooded in January and starved by March.
1. Measure completion, not occupancy. Put throughput on the wall next to status. How many initiatives finished this quarter? What is the average time from start to done, and is it falling? A board that watches those two numbers will make different decisions from one that watches a field of green, because it is finally looking at the thing that pays.

None of this requires a new methodology or a heavier office. It requires the reverse: the restraint to keep fewer things open, and the courage to defend the empty slot against the perfectly reasonable initiative that wants to fill it.

10. The courage of the empty slot

The hardest thing to hold, in the end, is the empty slot — the deliberately unspent capacity, the initiative left in the register rather than the active list, the yes withheld not because the idea is poor but because the organisation is full. Every instinct and every incentive argues against it. An empty slot looks like waste. A withheld approval looks like timidity. A shorter active list looks like a change function that has lost its ambition.

But the empty slot is where flow lives. It is the slack that lets the nearest-complete initiative sprint to done instead of waiting behind a queue. It is the honesty that says demand exceeds capacity and refuses to pretend otherwise by loading the surplus into a pipeline where it will only rot. The organisations that deliver most are not the ones with the fullest walls. They are the ones disciplined enough to keep the wall thinner than they comfortably could, so that what is on it actually moves.

We reward the full pipeline because it flatters us — it says we are ambitious, trusted, capacious. The uncomfortable truth is that ambition is demonstrated by what finishes, not by what is approved, and the surest way to finish more is to start less. The pipeline illusion is the belief that a portfolio's fullness is a measure of its strength. It is, far more often, a measure of how much we have quietly agreed not to deliver.